**AFRL-RI-RS-TM-2007-15**
**In-House Technical Memorandum**
**December 2007**


# MULTIMEDIA INFORMATION EXTRACTION

*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

**STINFO COPY**

**AIR FORCE RESEARCH LABORATORY**
**INFORMATION DIRECTORATE**
**ROME RESEARCH SITE**
**ROME, NEW YORK**

# NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TM-2007-15 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/ /s/

CHARLES G. MESSENGER
Chief, Situation Awareness Branch

JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

# REPORT DOCUMENTATION PAGE

*Form Approved*
**OMB No. 0704-0188**



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| DEC 2007 | | Jul 06 – Sep 07 |

**4. TITLE AND SUBTITLE**

MULTIMEDIA INFORMATION EXTRACTION

**5a. CONTRACT NUMBER**
In-House

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**
N/A

**6. AUTHOR(S)**

Emily Budlong

**5d. PROJECT NUMBER**
N/A

**5e. TASK NUMBER**
N/A

**5f. WORK UNIT NUMBER**
N/A

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
AFRL/RIED
525 Brooks Rd
Rome NY 13441-4505

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

AFRL/RIED
525 Brooks Rd
Rome NY 13441-4505

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSORING/MONITORING AGENCY REPORT NUMBER**
AFRL-RI-RS-TM-2007-15

**12. DISTRIBUTION AVAILABILITY STATEMENT**
*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# WPAFB 07-0539*

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
This paper describes the need for information extraction technologies within the military, some of the current technologies available, and the problems associated with them. It also looks at some of the ongoing research projects in areas of multimedia information extraction. Finally, it looks at the StreamSage audio extraction software and the demonstration of this software, explains how to run the software and the demo, and evaluates them.

**15. SUBJECT TERMS**
Multimedia, information extraction, audio, image, video

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Sharon M. Walter |
| U | U | U | UL | 12 | 19b. TELEPHONE NUMBER (*Include area code*) N/A |

# Multimedia Information Extraction


## Abstract

As stated at the 2003 Workshop on Satellite Data Applications and Information Extraction, there is a pressing need in today's world to be able to "facilitate quick war fighting decisions that fully leverage the huge volumes of available information." However, this can be a daunting task due to the massive amounts of data used to find intelligence about current situations and to predict future events. Information extraction technologies are currently being developed to aid intelligence analysts in the process of finding pieces of key information in text, audio, image, and video documents. This technology will have the ability to automatically pull out relevant pieces of intelligence and structure them in a way that analysts can easily understand, enabling them to utilize more data, make decisions faster, and stay on top of global situations.

This paper describes the need for information extraction technologies within the military, some of the current technologies available, and the problems associated with them. It also looks at some of the ongoing research projects in areas of multimedia information extraction. Finally, it looks at the StreamSage audio extraction software and the demonstration of this software, explains how to run the original software and critiques it, and describes the demo platform developed during the author's summer employment at AFRL-Rome.


## Background

Intelligence analysts must sort through large amounts of data to determine which data sources are relevant to their needs and to find pertinent pieces of information within that data. With the heightened security threats in today's world, there is an increased need to identify possible threats quickly and more accurately. However, the volume of data used to find intelligence is increasing while the number of intelligence analysts is decreasing due to downsizing within the Department of Defense. Therefore, the need for faster methods of searching intelligence data is growing. Information extraction (IE) is the process of extracting data from retrieved documents and saving it in a useful manner, such as a database. In the future, the use of IE technology will reduce the amount of time and labor spent on finding key pieces of information, therefore enabling analysts to develop reports faster and find relevant information in a shorter amount of time.

Text information extraction technology is one area of IE technology. It is divided into three levels of complexity: shallow, intermediate, and deep extraction. Current work in text IE has been at the shallow and intermediate levels. Shallow extraction finds basic information such as named entities (the names of people, places, and organizations), numerical information (monetary values, percentages, etc), and simple events denoted by action verbs. Technology in this area is currently available for use in the intelligence industry, thereby giving analysts the ability to sort through large amounts of text documents quickly to find basic information. Intermediate extraction technologies possess the capabilities to extract relationships between entities (e.g. Person B works for Person A) and find more meaningful event information, such as people taking part in an action, as well as the time and place of an action. This area is currently being researched by various organizations, including the United States Air Force Research Laboratory Rome Research Site. Future research in text IE will be at the deep extraction level. Deep extraction will recognize event scenarios and complex relationships, as well as infer information from non-explicit events. Deep extraction technology will also find information

pertaining to a specific entity across collections of documents and consolidate the information for easy access.

**Current State-of-the-Art**

Besides text extraction, areas of multimedia information extraction include audio, image, and video IE. Audio IE technology could be used to obtain information from conversations, radio and news broadcasts, and lectures. It will be useful within the intelligence community since most of the information transferred between people is through speech. Current audio IE technology uses automatic speech recognition (ASR) to generate a transcript of the audio file, and then finds information in the transcript. ASR is the process of using a computer algorithm to convert speech to a set of words that can be written as text. Once the transcript is generated, programs can search the text for pieces of key information, such as named entities, similar to the process of text IE. However, accuracy can become a problem here due to errors in transcription, such as the insertion, deletion, and substitution of words during the conversion from speech to text. Research is currently being done to improve the accuracy of ASR. Work is also being done in the field of multilingual audio IE. This technology will be of great value to the military as it has the ability to perform IE in a foreign language and then convert the information into the user's native language, thereby reducing the need for a translator.

Image IE technology will be used to search large volumes of images to find images relevant to a query and recognize some of the objects within the image. Some military uses of this technology are to: plan travel routes and estimate travel time; find obstacles, landmarks, and shortcuts; and to find potentially dangerous spots along possible travel paths. Currently, image IE uses keywords and content-based retrieval (CBR) to find images relevant to queries. CBR is the process of searching a database using the contents of the desired multimedia, rather than keywords or captions. It also has the ability to detect the low-level features of an image, such as recognizing an orange and black blob in a picture of a tiger. In addition, CBR can be used to find images based on their similarity to other images. This is particularly useful when searching for more images of the same person or object. Current research is investigating more advanced levels of CBR to enhance search capabilities and accuracy.

Uses of video IE technology include planning travel routes and finding trouble spots, finding specific people in videos, and identifying abnormalities in security videos such as a person crossing railroad tracks in an undesignated area. Current technology in this area can segment video streams into sections based on scene changes, speaker changes, or other user-defined criteria. It can also recognize different objects (e.g. a person versus an animal) and detect when a scene changes from its normal set-up. Video IE technology frequently makes use of audio IE technology, especially when segmenting video streams by speaker changes or topic changes, by recognizing pauses in sound or changes in voice. Research in this area is currently exploring how to accurately recognize people of interest within a video.

**Problems**

One problem currently being encountered in information extraction technology research is handling the move from text to multimedia extraction. A prominent problem in this shift is how to index the masses of multimedia data. Documents need to be indexed for fast retrieval; however, there is no universal, generic way of describing and retrieving information of varying media that is acceptable. Unlike text documents, most of which have an index that directs the reader to a specific area to find desired information, there is no single method to tell a user where

to go to find specific information within multimedia. Creating an indexing system for multimedia data is difficult because manual indexing is subjective and the purpose of indexing the data frequently differs from the reason for retrieval. This means that there is no standard way of describing the elements of the data and that the data can have different descriptions and interpreted meanings depending on the reasons for its use. In addition, text descriptions of multimedia data are too subjective to use for indexing due to the fact that there is no single way to describe it for retrieval.

Feature extraction extracts various features from an image (e.g. texture, color, shape) to identify and interpret meaningful physical objects within an image. Feature extraction offers a more objective method of indexing, however, it is insufficient for retrieval as information retrieval (IR) is subjective and based on humans' notions about the data. These problems with creating objective indices for databases are the reasons for the current lack of a universally acceptable indexing method. As a result, it is difficult to access a specific part of a database as access methods vary with different systems.

Multimedia semantics is the source of the indexing problems and also causes other problems with describing and retrieving data. There are many interpretations of any single piece of multimedia data as a result of the subjectivity of human nature. This inhibits the retrieval of data and also creates a need to manage the multiple meanings for the same material to facilitate information retrieval. However, a single model to facilitate IR based on user-defined semantics does not currently exist. Active research in this area is trying to find more objective and universally acceptable ways to describe and index multimedia to facilitate IR.

Another problem within IE technology is the "semantic gap." The semantic gap refers to the difference between identifying the low level features of an object, such as its color, shape, and texture, and identifying the object correctly. Currently, a single low level description maps to several different objects matching that description. For example, searching for an apple using the description "red, smooth, and somewhat spherical" can result in several different images fitting that description being returned (see Figure 1). Researchers are currently addressing this issue to find a way to specify the low level description in a way that the correct image is returned.

**Figure 1**

CBR Query: red, smooth, somewhat spherical object
Results: (images are a red ball, an apple, a tomato, and a red bean bag chair)

**Some Current Research Projects**

The Speech and Information Extraction for Video Exploitation (SIEVE) program is a research project of SRI International funded by the Defense Advanced Research Project Agency (DARPA) in the area of audio information extraction. SIEVE will be a sophisticated "news on demand" system, allowing users to find news segments on topics they are interested in within a large volume of news broadcasts. One goal of the research is to segment the files by speaker through the identification of recurring speakers, segment files using linguistics by detecting sentence boundaries and disfluencies and corrections in speech, and segment files by topic using pitch changes and pauses in speech. SIEVE will also have IE capabilities and be able to find names within audio files. The project is using knowledge source integration, which is the process of combining existing recognition technologies, in order to expand and enhance the capabilities of current technologies.

Image IE technology research is being continued in the GNU Image Finding Tool (GIFT). GIFT is a research project of the vision group at the computer science center of the University of Geneva. GIFT uses content-based image retrieval (CBIR) to search through volumes of images using the content of images and query by example (QBE). QBE is the process of searching an image database using an example image as a query. The program also makes use of relevance feedback to improve search results. Relevance feedback is the use of relevant results from an old query to perform a new query. There is no need to annotate images when using GIFT, as the program does not use keywords to search for images; it only searches the content of the images. GIFT also has the ability to index image directory trees, making it easier and faster to find images that may be related to each other.

Carnegie Mellon University's Extensible News Video Information Exploitation (ENVIE) project, funded by the Advanced Research and Development Agency (ARDA) Video Analysis and Content Exploitation (VACE) program, is continuing research within the video IE area. ENVIE has the capability to automatically detect, extract, and report people, patterns, and trends of interest from the visual content of domestic and foreign news broadcasts. It can also derive comprehensive video events through the identification of people and object relationships over time and location. Essentially, the program investigates, analyzes, and summarizes news content according to user-defined analysis criteria. ENVIE does this by classifying video sequences and audio features to improve its interpretation of news events. The program also allows the analyst to mark and add notes to video shots, as well as skim through videos to quickly find sections of interest.

**StreamSage Audio Extraction Software**

AFRL-Rome worked with StreamSage, Inc. to investigate audio extraction technology through the Air Force Dual-Use Science and Technology Program within which the contractor, the Air Force, and AFRL shared project expenses. The project was a first foray by AFRL-Rome into investigation of various extraction media, toward development of a concerted research program in multimedia information extraction for military application.

The StreamSage software creates a transcript of an audio file then searches the transcription for the existence of phrases entered as a query by the user, similar to how a text search engine works. It allows a user to enter queries and then view the results as a list of audio files that mention the query within it. The user can also see keywords for each audio file in the results, view a speech excerpt relevant to the query, delete queries, and listen to the relevant audio section or the entire audio file. The software also allows users to clear their profile and

access information about the program. It also saves the user's profile until it is explicitly told to clear it.

To run the StreamSage software, begin by turning on the computer and logging in. Once the computer has finished starting up and is displaying the desktop, double click on the StreamSage icon. This will run a batch file. When it is done, open Internet Explorer and enter http://localhost:8080/afir in the address bar and hit the enter button on the keyboard. This will finish opening the software and allow the user to begin.

To start using the software, enter a query (e.g. "Congress") in the text box and press the "Add" button. This will refresh the window and display the results under "User Entity Results". Keep doing this as desired.

The "Entity Watch List" displays each entity the user has queried, the option to delete it, and the number of news and conversation files that were found as results. Clicking on an entity in this list will display the results of that query under "User Entity Results." Clicking "Delete" will delete the entity from the "Entity Watch List," as well as the results. The user can also view the top two results for every entity by clicking on "Show Files For All Entities" at the bottom of the "Entity Watch List."

The "User Entity Results" section displays the results for each query. Here a user can play the entire audio file or just the relevant segment, view the speech excerpt, see the keywords for the file, or rate the file. To play the entire file, the user should click on the audio file name to the left of the italicized keywords. To play the relevant section, the user should click on "Play Relevant Section" for the result he wants to hear. Both actions will open Windows Media Player and automatically begin playing the file; the user can control the playback using the controls in the media player.

To view the speech excerpt, hover the mouse over "Speech Excerpt" and a pop-up will display a transcript of the relevant audio section. Clicking on "Speech Excerpt" will only display a page that reads "TEST." Clicking on "Rate Relevant File" will open a new window where the user can select a relevance level for the file. However, clicking the "Rate" button in this window only resets the window, it does not change the rating of the file. At any point when running this software, the user may press the "Clear" button to delete all entities and results and clear their profile. The user may also click the question mark to view information about the program.

To exit the program, exit Internet Explorer by clicking the X in the upper right corner and close the command prompt window in the same manner.

This program does not do audio IE directly from the audio file; however it does match queries to phrases in audio files. For example, a search for "Congress" will return all files that contain the word "Congress" in it. The software is relatively easy to use, although it can be frustrating at times since it appears to have some functionality that it really doesn't (e.g. rating file relevancy). The speech excerpts of each file are relatively accurate and give the user an idea of what is being said, although if the excerpt is long the user may not be able to see the entire text. It also consolidates all relevant audio sections into one file to play using "Play Relevant Sections", which is convenient when there are multiple sections in the same file that are relevant to the query. The program also uses a well-known media player (Windows Media Player) to play files which eliminates the need to learn how to work with a new media player. However, the keywords shown for files are not always helpful as they frequently are inaccurate for the content of the audio file. In addition, the software is difficult to install and does not appear to allow the user to add their own audio files into the system. Overall, the software is a good example of technology capable of finding entities within an audio file, returning the results to the

user, and allowing the user to access the results easily, but is not really suitable for use in the intelligence community due to its inaccuracies in transcribing audio files, lack of relevance feedback, and lack of true IE capabilities.

**Audio Project and Demo**

The StreamSage Demo is a software demonstration of the StreamSage audio extraction software developed by the author during summer employment at AFRL-Rome. The demo system has most of the functionality of the StreamSage software, but operates at a lower level. Similar to the StreamSage software, the StreamSage Demo does not extract information directly from the audio, but rather searches the transcripts of the audio files to find phrases matching the user's input. Its functionalities allow users to enter queries and then the program returns the results of that query. The program also allows users to view speech excerpts, play audio files, rate audio files, sort the results by file name or relevance, and view keywords for each audio file returned as query results. Unlike the original software, the StreamSage Demo also allows the user to add their own audio files into the system.

The StreamSage Demo includes a new, simpler interface that was developed to facilitate demonstrations of the software. The demo is also designed so that it can quickly and easily be installed on different computers.

To run the StreamSage Demo, turn on the computer and log-in. Once the desktop is showing, double click on the RunSSDemo icon; this will start the demo. Once the window appears, begin by entering queries into the text box at the top and pressing "Search and Add." This will cause the program to search the audio file transcripts for phrases matching the query, and then return the corresponding audio files as a list of results in the "Files" section of the window and add the query to the "Entity List." To view results for an entity, select the desired entity in the "Entity List" and then press the "View Results" button in the "Options" panel. To delete an entity, select the entity and click the "Delete" button. To work with the results, begin by selecting a file in the "Files" panel. Selecting a file will display the keywords and the file's relevance to the query in the "Keywords and Relevance" area. Once a file has been selected, the user may click the "Speech Excerpt," "Play File," or "Rate File" button. Selecting "Speech Excerpt" will open a new window with an excerpt of the transcript in it; to exit this window click the X in the upper right corner of the excerpt window. Clicking "Play File" will open a window with a controller in it and automatically start playing the file. The user can pause the playback by pressing the pause button on the controller, or skim through the file by moving the circle along the bar. To completely stop the playback, exit the window by clicking the X in the upper right corner of the playback window. Pressing the "Rate File" button will open a new window where the user can select a rating for the file by clicking in one of the three radio buttons and pressing OK, or the user can click the X in the upper right corner to exit the window and not rate the file. At any point during the execution of the program, the user may view information about the program by going to the Help menu and selecting About. The user may also clear all searches by pressing the "Clear" button, or he can view the top three results for all searches by pressing the "Show Results For All Entities" button. In addition, the user can change how the results are sorted by going to the File menu and selecting Sort by File Name or Sort by Relevance. The default setting is to sort the results by file name. The change in sorting will not be visible until the user refreshes the "Files" panel by pressing the "View Results" button again. The user should also note that the "View Results" and "Delete" buttons will be enabled only when an entity is selected, and that the "Speech Excerpt," "Play File," and "Rate File" buttons

will be enabled only when a file is selected. Also, the status bar at the bottom of the window displays information about what the program is doing, such as playing a file, and information about errors, such as trying to query duplicate entities, throughout the running of the program. To exit the program, press the X in the upper right corner of the screen, go to the File menu and select Quit, or close the command prompt window.

This demo is a simple example of finding specific information in audio files. It searches transcripts of audio files for the existence of specific words or phrases and then returns the corresponding audio files to the user. It is fairly simple to use and handles user errors well. It is also informative as to what it is doing, lets the user know when they have tried to do something they should not have, and also ensures that the user is only capable of executing functions applicable to their selection (entity versus file). The program can also do some multitasking. For example, the user can listen to an audio file while performing other operations, such as viewing the speech excerpt, viewing other files, or performing new queries. The demo system is also easy to install on different machines. However, the transcripts generated for this program are very inaccurate due to the high number of word substitutions, insertions, and deletions. In addition, they frequently make no sense when trying to read through them because of these inaccuracies. This demonstrates the errors that can occur during the transcription of audio files and the effect they can have on audio information extraction. Therefore, this demo is not suitable for use other than as an example of the StreamSage software capabilities. This demo also requires extra work before running the program if the user wishes to change the audio files used at any point; the program requires that automatic speech recognition be run before attempting to use the program, as it needs pre-generated transcript files to run. It also can be slightly cumbersome, particularly when changing how the results are sorted, as the program does not automatically refresh everything whenever a setting changes, forcing the user to manually refresh any affected parts. Overall, the StreamSage Demo works as a software demonstration as expected; it demonstrates most of the functionality of the StreamSage software and is very portable, but operates at a lower, less accurate, and less user-friendly level.

**Summary**

During the course of the Dual-Use Science and Technology project that developed the original software investigated in this effort, StreamSage, Inc. was bought out by ComCast and no longer exists. ComCast does not appear to be continuing support and development of this software. Other companies have, of course, continued audio extraction research.

State-of-the-art IE technologies are starting to become useful in intelligence analysis and are showing promise for extended use in the future. Research is continuing to resolve the problems currently being encountered, such as the problems with indexing and retrieving multimedia documents, and improving current technologies. Work must be done to make IE technologies robust so that they can handle unexpected and erroneous inputs and to make them accurate so they are reliable for military and commercial use. Research with referencing the same entity across multiple documents needs to continue, as well as research in multilingual IE technologies to make the technology more useful in intelligence applications. In addition, advanced IE capabilities must also be extended beyond text to multimedia documents in order to be able to fully leverage the masses of data available for analysis.

This project was a first foray into investigation of various extraction media, toward development of a concerted research program in multimedia information extraction for military applications.

**Glossary**

ARDA – Advanced Research and Development Agency

automatic speech recognition (ASR) – process of using an algorithm implemented as a computer program to convert a speech signal to a set of words

content based image retrieval (CBIR) – process of searching a database of images using content examples rather than keywords or captions

content based retrieval (CBR) – process of searching a database using contents of desired multimedia rather than keywords or captions

DARPA – Defense Advanced Research Project Agency

ENVIE – Extensible News Video Information Exploitation

feature extraction – extracting various image features to identify and interpret meaningful physical objects within the image

GIFT – GNU Image Finding Tool

indexing – automatically describing documents to facilitate quick retrieval

information extraction (IE) – process of extracting data from retrieved documents and saving it in a manner that makes it more useful

information retrieval (IR) – searching for and retrieving documents in response to queries for information

knowledge source integration – process of combining existing recognition technologies to develop better technologies

query by example (QBE) – search documents using example documents as a query

relevance feedback – use of information about which results are relevant to perform a new query

semantic gap – difference between identifying the low level features of an object and identifying the object

SIEVE – Speech and Information Extraction for Video Exploitation

temporal changes – changes taking place over time

VACE – Video Analysis and Content Exploitation

**References**

Betz, Chris. "GIS and Multimedia in the Military." March 18, 2003. http://www.directionsmag.com/article.php?article_id=306&trv=1, 2/06.

"ENVIE: Extensible News Video Information Exploitation." http://www.informedia.cs.cmu.edu/arda/vaceII.html, 2/06.

"The GNU Image Finding Tool." http://www.gnu.org/software/gift, 7/06.

Pine, Carrie. "Information Extraction (IE) Technology," unpublished paper.

Schouten, Ben. "The Semantic Gap: Words and Images." http://framework.v2.nl/archive/archive/node/text/.xslt/nodenr-136433, 7/06.

"Speech and Language Technology for Information Access and Processing." http://www.speech.sri.com/projects/sieve, 6/06.

Srinivasan, Uma and Surya Nepal. Managing Multimedia Semantics. IRM Press, Hershey, London, Melbourne, Singapore, 2005.

Wactlar, Howard et al. "ENVIE: Extensible News Video Exploitation for Intelligence Analysis." https://analysis.mitre.org/proceedings/Final_Papers_Files/147_Camera_Ready_Paper.pdf , 6/06.

Wade, Gary S. and Jeffrey R. Key. "Workshop on Satellite Data Applications and Information Extraction." November 25, 2003. http://stratus.ssec.wisc.edu/meetings/data_ws/data_ws_report.pdf, 6/06.

Walter, Sharon M., Carrie Pine, and Walter Gadz. "Information Extraction Technology at the US Air Force Research Laboratory Rome Research Site (AFRL-RRS)," unpublished paper.